aps
ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Computational Models of Face Perception

Current Directions in Psychological Science
1–7

DOI: 10.1177/0963721417698535
www.psychologicalscience.org/CDPS
SAGE

Aleix M. Martinez
Department of Electrical and Computer Engineering, Center for Cognitive and Brain Sciences, and Mathematical Biosciences Institute, The Ohio State University

Abstract
Faces are one of the most important means of communication for humans. For example, a short glance at a person's face provides information about his or her identity and emotional state. What are the computations the brain uses to acquire this information so accurately and seemingly effortlessly? This article summarizes current research on computational modeling, a technique used to answer this question. Specifically, my research tests the hypothesis that this algorithm is tasked with solving the *inverse problem* of production. For example, to recognize identity, our brain needs to identify shape and shading features that are invariant to facial expression, pose, and illumination. Similarly, to recognize emotion, the brain needs to identify shape and shading features that are invariant to identity, pose, and illumination. If one defines the physics equations that render an image under different identities, expressions, poses, and illuminations, then gaining invariance to these factors can be readily resolved by computing the inverse of this rendering function. I describe our current understanding of the algorithms used by our brains to resolve this inverse problem. I also discuss how these results are driving research in computer vision to design computer systems that are as accurate, robust, and efficient as humans.

Keywords
face recognition, face processing, affect, categorization, language evolution

After finishing this sentence, look around, find a person you know, and then look briefly at his or her face. What can you tell about this person? Likely the person's name (identity) and emotional state come to mind. Most of us effortlessly extract this information from the smallest of glimpses.

This article reviews our current understanding of the computations that are performed by the brain to achieve these seemingly effortless tasks—visual recognition of identity and emotion. We assume that the brain is a type of computer running algorithms specifically dedicated to the interpretation of other people's faces. The goal is to decode and understand these algorithms using computational modeling. Specifically, this article details how current progress in computational modeling is helping us understand how the brain recognizes faces.

My use of computational models is based on the hypothesis that the brain is tasked with solving the *inverse problem* of image production. That is, if f(.) defines how a facial attribute maps onto an image in the retina, then the brain's goal is to solve the inverse problem, $f^{-1}(.)$—how the image on one's retina translates into understanding a facial attribute.

For example, imagine you are looking at Sally's face. Here, the brain's goal is to recover the name "Sally." More formally, the retinal image, $\boldsymbol{I}$, is equal to Sally's face: $\boldsymbol{I}$ = f(Sally's face). And the goal is to compute the inverse function: Sally's face = $f^{-1}(\boldsymbol{I})$.

The identity of someone's face is engraved in the person's three-dimensional face structure and the reflectance properties of the person's skin. These are examples of some of its diagnostic features. But this is not what we see. Rather, the two-dimensional shape of the face on your retinal image depends on the viewing angle and the person's facial expression. The brain's goal is to uncover the diagnostic features and filter out variations due to expression, viewing angle, and illumination.

It is imperative to note that computational modeling is useful only if it identifies these diagnostic features, algorithms, and mechanisms involved in the recognition of faces. In the following sections, I show that some machine-learning approaches, such as deep learning, are not generally helpful for answering these questions.

Corresponding Author:
Aleix M Martinez, 205 Dreese Laboratories, 2015 Neil Ave., The Ohio State University, Columbus, OH 43210
E-mail: martinez.158@osu.edu

Fig. 1. An image of a face (left) and its corresponding shape (right). (Face image drawn from Du, Tao, & Martinez, 2014.)

Recognition of Identity

Look at the left image in Figure 1. This is a two-dimensional image, $\boldsymbol{I}$. Now, look at the image to its right. This image defines the shape, $\boldsymbol{s}$, of the main components of that face.

Given many face images and their shapes ($\{\boldsymbol{I}_i, \boldsymbol{s}_i\}$, $i = 1, \ldots n$), one can compute the mean shape as well as the major differences (i.e., the largest standard deviations) between shapes. These variances are given by principal component analysis, a statistical technique that allows us to find the shape features that produce maximum variability (Martinez & Kak, 2001). The resulting representation is called a *norm-based face space,* because all faces are described as deviations from the mean (norm) sample shape (Leopold, Bondar, & Giese, 2006), Figure 2.

Recall, however, that our retinal image, $\boldsymbol{I}_i$, is two-dimensional but that diagnostic features exist in three-dimensional space. Can we design an algorithm that estimates the three-dimensional shape of a face from a single image? Yes. In fact, everyday experience proves this. When you looked at a face at the beginning of this article, you probably just saw it from a single viewing angle. Yet you were able to mentally imagine other views of that face as well.

My students and I have shown that the computations needed to solve this problem are quite simple (Zhao, Wang, Benitez-Quiroz, Liu, & Martinez, 2016). The algorithm works as follows: Given a set of two-dimensional images and their corresponding three-dimensional shapes ($\{\boldsymbol{I}_i, \boldsymbol{S}_i\}$, $i = 1, \ldots, n$), we use a machine-learning technique called regression (You, Benitez-Quiroz, & Martinez, 2014) to learn the functional mapping from a retinal image to a three-dimensional shape, $\boldsymbol{S}_i = \mathrm{f}(\boldsymbol{I}_i)$; see Figure 3 for an illustration of such a regression. Once this function has been learned using the available training data, we can use it to estimate the three-dimensional shape, $\hat{\boldsymbol{S}}$, of a previously unseen face image, $\hat{\boldsymbol{I}}$, as expressed in the following equation: $\hat{\boldsymbol{S}} = \mathrm{f}(\hat{\boldsymbol{I}})$.

The model above allows us to map an initial face image to a rotation-invariant representation. Physiological studies, though, suggest the existence of an intermediate

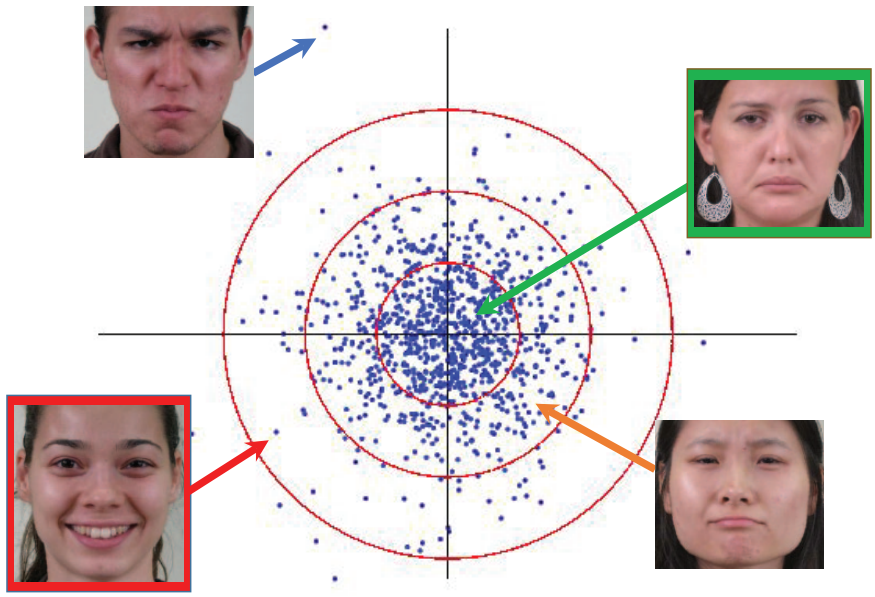

Fig. 2. A norm-based face space. Shown here are the two dimensions with largest variance in the shape space; the four face images correspond to each of the four feature vectors, respectively. The farther away a face is from the origin of this space, the easier it is to recognize. Here, the face demarked with a red border (bottom left) is easier to recognize than the face delineated with a green border (top right). (Face images drawn from Du, Tao, & Martinez, 2014.)

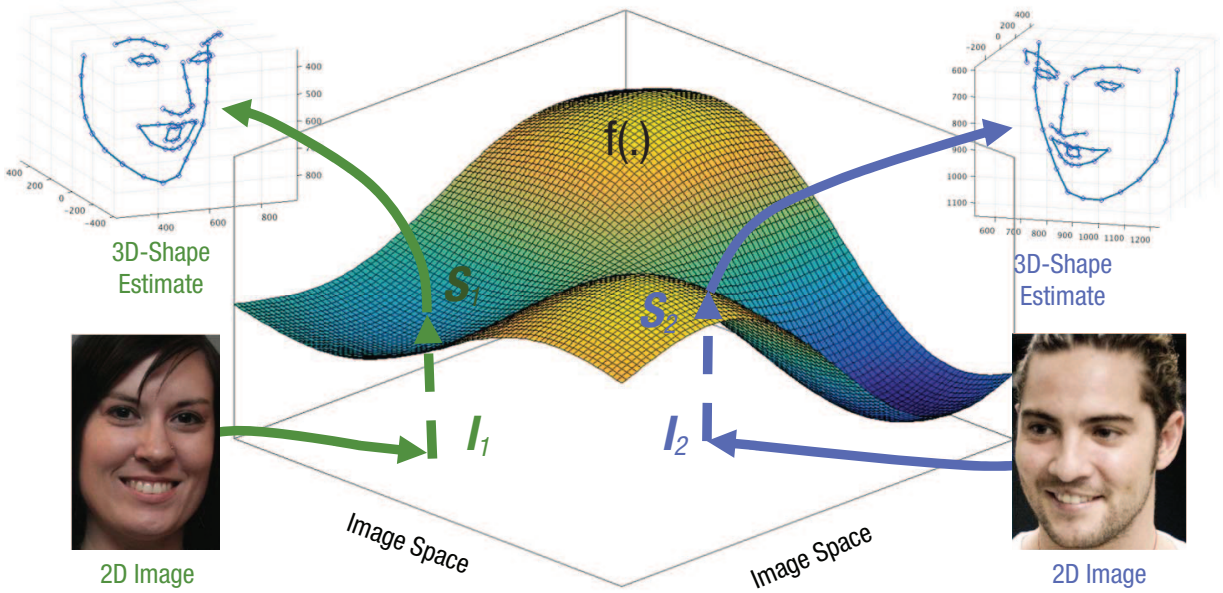

Fig. 3. A plot illustrating a regressor estimating the three-dimensional shape of a face from a two-dimensional image. Here, two of the axes define the image space. The third (vertical) axis defines the three-dimensional shape. Of course, in reality, the image space and the 3D-shape space are defined by more dimensions. The function f(.) is the regressor, which defines a non-linear manifold. This manifold specifies the mapping between an image, $\boldsymbol{I}_i$, and its three-dimensional shape, $\boldsymbol{S}_i$: $\boldsymbol{S}_i = f(\boldsymbol{I}_i)$. (Face images drawn from Du, Tao, & Martinez, 2014.)

representation invariant to mirror images of profile views (Meyers, Borzello, Freiwald, & Tsao, 2015). A simple analysis based on projective geometry shows that the basis functions (e.g., principal components) obtained with the above formulation yield the same response for mirror-symmetric images (Leibo, Liao, Anselmi, Freiwald, & Poggio, 2016). As an intuitive proof, note that faces are symmetric about the vertical midline of the face. Thus, rotating a face 90° to the left and 90° to the right yields basically the same image, up to a mirror projection.

Deep Learning

If the function f(.) presented above is defined by many parameters, the resulting regression approach is called *deep learning*. In deep learning, one must use a very large number of training samples to successfully estimate that same number of parameters, an increasingly popular approach called *big data*.

Deep learning has recently achieved good results in addressing several problems in computer vision, including face recognition (e.g., Kemelmacher-Shlizerman, Seitz, Miller, & Brossard, 2016). Unfortunately, this technique does not generally help us uncover the underlying computations of the algorithm used by our visual system. For one, we do not yet know how to apply deep learning to solve many problems in face recognition—for instance, recognition under varying illumination, or the recognition of emotion. Also, the reliance on big data makes deep learning an unlikely model of human vision. Humans generally learn from a single sample (Martinez, 2002, in press), not the thousands required by current deep-learning algorithms. And, crucially, deep learning does not generally provide information on the mechanisms used by the brain to decode facial attributes. That is, we might be able to design computer algorithms that identify people's faces very accurately, yet learn nothing about the brain.

To clarify this point, imagine a physicist trying to understand the behavior of a number of particles. Given enough observations of the behavior of these particles, deep learning could certainty be used to identify a function describing their behavior. This function would allow us to predict the state, y, of the particles, x, after an event, f(.): $y = f(x)$. But this would not provide any insights into the mechanisms involved in that process—that is, the laws of physics. The same applies to the study of the visual system. It is not sufficient to demonstrate that there exists a function that maps an image to a facial attribute. We also wish to uncover the specific computations used by the brain to accomplish this. I argue that we need to refocus our research toward computational models that can solve this problem.

Gilad, Meng, and Sinha (2009) suggested that local contrast polarities between a few regions of the face (especially those around the eyes) encode critical information about a person's identity and that the brain uses this information to recognize people's faces. Subsequently, Ohayon,

Fig. 4. The *American Gothic* illusion. The image on the left is the male character in Grant Wood's famous *American Gothic* painting, typically described as having a sad expression. However, this man is not expressing any emotion; if you look closely, you will see that his face is at rest, assuming a neutral expression. Research suggests that the character appears to be sad because Wood painted him with an elongated face (i.e., a very thin face), including an exaggeratedly long distance between his brows and mouth. The image on the right is a morphed image, manipulated to be wider and to have a significantly shorter brow-to-mouth distance—changes that cause the character to be perceived as angrier. These results are consistent with the predictions of our computational model.

Freiwald, and Tsao (2012) identified cells in the macaque monkey brain that selectively respond to such contrast variations. Computer vision algorithms based on this local contrast polarity successfully detect and recognize faces in images (Turk, 2013; Zhao et al., 2016). Furthermore, this model explains how one is able to recognize partially occluded and expression-variant faces (Jia & Martinez, 2009)—for example, using graph matching (Aflalo, Bronstein, & Kimmel, 2015; Zhao & Martinez, 2016). These results do point toward an understanding of some of the computations underlying the perception of faces.

Facial Expressions of Emotion

Another aspect of face perception is our remarkable ability to interpret facial expressions. Facial expressions convey a lot of information, such as a person's emotional state. Like the representation of face identity, the representation of facial expression uses a norm-based model (Neth & Martinez, 2009, 2010). However, the dimensions employed in the recognition of emotion are, for the most part, different (Martinez & Du, 2012; Richoz, Jack, Garrod, Schyns, & Caldara, 2015; Sormaz, Young, & Andrews, 2016). We therefore say that the form of the face space is the same for expression and identity but that the dimensions defining this space differ between the two. What are the features that represent these dimensions, then?

Studying the physical reality of facial expressions shows that they are produced by contracting and relaxing different muscles in the face (Duchenne, 1862/1990). Thus, I hypothesize that the brain solves the inverse problem by attempting to decode which facial muscle actions, $h(\boldsymbol{I})$, are active during a particular expression (Martinez, in press). My research group has recently developed a computer vision system based on this model, using an algorithm that accounts for shape and shading features and incorporating it into machine-learning algorithms that identify which of these features best discriminate the muscles involved in each expression (Benitez-Quiroz, Srinivasan, & Martinez, 2016; Du, Tao, & Martinez, 2014).

Take the example of a small cheek muscle that is used to pull the lips outward to create a smile. Unsurprisingly, our machine-learning approach identified shape and shading changes in the corners of the mouth as the most discriminant feature for smiles. Likewise, contracting a set of three facial muscles located at the top of the face results in the lowering of the inner corners of the brows. Yet the most discriminant shape and shading features that allow us to detect this facial action are associated with more distal parts of the face—the brow-to-mouth distance and the face's height-to-width ratio—because they change when one contracts these muscles (Du et al., 2014). Accordingly, the algorithm assumes that these muscles are active when processing the faces of people who have very thin faces with unusually large distances between their brows and mouths (Martinez, in press).

This effect is clearly visible in Figure 4. In the left image, we see the male character in Grant Wood's painting *American Gothic*. Note that this person is not expressing any emotion, yet you are likely to perceive sadness in his expression. Using morphing software, we can decrease the distance between his brows and mouth and make his face wider, as shown in the right image. Notice how the face now looks angry because we have incidentally altered it to display the image features associated with the facial muscle actions used to express anger.

If this algorithm is indeed implemented in our brains, then there should be an area of the brain dedicated to the detection of these facial muscle actions. In a recent article, my research group identified one such region just behind the right ear: the posterior superior temporal sulcus (Srinivasan, Golomb, & Martinez, 2016).

Compound Emotions

An ongoing debate in emotion theory concerns the number of facial expressions that people can visually recognize. Darwin (1872/1965) argued that six facial expressions of emotion can be visually recognized across cultures. However, my group's computational modeling presented above

Fig. 5. The "not face." This expression is used as a marker of negation in at least four different languages—that is, speakers may produce this expression when they want to convey a negative statement (e.g., "No, I didn't go to the party"). In sentences in American Sign Language (ASL), this expression may be the sole marker of negation—a grammatical marker. The images, from left to right, show the "not face" as expressed by native speakers of Mandarin Chinese, Spanish, and ASL. (Face images drawn from Benitez-Quiroz, Wilbur, & Martinez, 2016.)

suggests that the visual system does not attempt to categorize facial expressions but, rather, simply identifies the facial muscle actions involved in the production of expressions. Why should that be, given that it is obviously easier to visually identify six facial expressions than to try to decode individual facial muscle actions? My hypothesis is that by identifying facial muscle actions, the visual system can categorize many more than six facial expressions. Our current model suggests that people might be able to recognize over a hundred categories of expressions (Benitez-Quiroz, Srinivasan, & Martinez, 2016). It is certainly easier to identify a few facial muscle actions than to derive an algorithm that can discern such a large number of categories.

So far, we have identified 23 facial expressions of emotion, including compound emotions (e.g., angry surprise, happy disgust; Du & Martinez, 2015; Du et al., 2014). We are currently studying an even larger number of facial expressions that correspond to about 400 affect concepts (e.g., anxiety, embarrassment, and fatigue; Benitez-Quiroz, Srinivasan, & Martinez, 2016). And, although we do not yet know which are universally used and recognized, our preliminary analysis suggests that the number of universally recognized expressions is much larger than current models propound.

Everyday experience seems to corroborate our ability to use the face to express many more than just a few emotion categories. People seem to use their faces to communicate a large number of concepts. But which ones?

Grammatical Markers

The ability to produce and visually recognize compound facial expressions of emotion allows people to communicate complex concepts nonverbally. Of note, I hypothesize that compound emotions have evolved into grammatical markers. For example, in a recent article (Benitez-Quiroz, Wilbur, & Martinez, 2016), my research group showed that compounding the facial expressions of anger, disgust, and contempt yields an expression that serves as a marker of negation. If this is part of human language, we can call it a *grammatical marker*—specifically, a grammatical marker of negation (i.e., negative polarity). This means you can use this expression to convert a positive sentence into a negative one.

We call this expression the "not face," and it is illustrated in Figure 5. We have shown that this compound facial expression of emotion is used in a variety of cultures and languages, including English, Spanish, Mandarin Chinese, and American Sign Language (ASL). Crucially, in ASL sentences, the "not face" is sometimes the sole marker of negation. That is, if you do not see the face of the signer, you may not know if the signed sentence is positive or negative.

A fundamental and unanswered question in the cognitive sciences is: Where does language come from? Whereas most of our human abilities can be traced back to similar or more primitive versions of the same ability in our closest living species, language cannot. The idea that the "not face" is a compound facial expression of emotion is significant because it implies a plausible evolutionary path for the emergence of language through the expression of emotion.

As significant as this implication might be, more research is needed to test this hypothesis. Uncovering the origins of language is one of the most exciting problems in science. But although the results discussed above have shown how computational models can aid in this search, additional studies will need to be completed to provide a clear picture of the emergence of grammatical markers through the expression of emotion.

How Many Facial Expressions Are There?

It is still unclear how many facial expressions are commonly used to communicate affect. Although research in my lab has provided strong evidence for the existence of many categories of expressions, other researchers have suggested that emotions are not represented categorically in the brain (Skerry & Saxe, 2015) and that their representation is not localized in a small brain area, as our results indicate (Wager et al., 2015). Others have argued for a hierarchical organization of emotions (Jack, Garrod, & Schyns, 2014): Given that facial expressions are dynamic, the hypothesis is that information conveyed earlier is more informative about a few important emotion categories, and later components of expressions are more social-specific.

Future research will hopefully resolve the details of the computations performed by our brains to interpret faces and facial expressions. This undertaking is important because the results will play a major role in the definition, diagnosis, and treatment of psychopathologies. At present, heterogeneity and reification of psychopathologies pose major challenges for translational research. It has been argued that a successful definition of the brains' functional impairments will require a detailed understanding of the brain's computational mechanisms (Insel, 2014; Insel et al., 2010). Computational models are ideally suited to address these problems.

Recommended Reading

Clark-Polner, E., Johnson, T. D., & Barrett, L. F. (2016). Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. *Cerebral Cortex*, *27*, 1944–1948. An accessible argument against the categorical model of six (basic) emotion categories.

Du, S., Tao, Y., & Martinez, A. M. (2014). (See References). An article that provides a computational model and analysis identifying a set of previously unknown facial expressions of emotion.

Phillips, P. J., & O'Toole, A. J. (2014). Comparison of human and computer performance across face recognition experiments. *Image and Vision Computing*, *32*, 74–85. An overview of recent face-recognition competitions in computer vision, with a comparison to human performance.

Sinha, P., Balas, B., Ostrovsky, Y., & Russell, R. (2006). Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, *94*, 1948–1962. A well-articulated, comprehensive overview of image transformations that do not affect our perception of faces but greatly impact computer vision systems.

Srinivasan, R., Golomb, J. D., & Martinez, A. M. (2016). (See References). An article that delineates the neural mechanisms that implement the computational model of the perception of facial expressions defined by the author's research group.

Acknowledgments

Special thanks to Fabian Benitez-Quiroz, Ramprakash Srinivasan, and Shichuan Du.

Declaration of Conflicting Interests

The author declared no conflicts of interest with respect to the authorship or the publication of this article.

Funding

Aleix M Martinez is supported in part by National Institutes of Health Grant R01-DC-014498, Human Frontier Science Program Grant RGP0036/2016, and The Ohio State University's Mathematical Biosciences Institute.

References

Aflalo, Y., Bronstein, A., & Kimmel, R. (2015). On convex relaxation of graph isomorphism. *Proceedings of the National Academy of Sciences, USA*, *112*, 2942–2947.

Benitez-Quiroz, C. F., Srinivasan, R., & Martinez, A. M. (2016). EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In *Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition (CVPR)*. Piscataway, NJ: IEEE.

Benitez-Quiroz, C. F., Wilbur, R. B., & Martinez, A. M. (2016). The not face: A grammaticalization of facial expressions of emotion. *Cognition*, *150*, 77–84.

Darwin, C. (1965). *The expression of the emotions in man and animals*. Chicago, IL: University of Chicago Press. (Original work published 1872)

Du, S., & Martinez, A. M. (2015). Compound facial expressions of emotion: From basic research to clinical applications. *Dialogues in Clinical Neuroscience*, *17*, 443–455.

Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences, USA*, *111*, E1454–E1462.

Duchenne, C. B. (1990). *The mechanism of human facial expression*. Cambridge, UK: Cambridge University Press. (Original work published 1862)

Gilad, S., Meng, M., & Sinha, P. (2009). Role of ordinal contrast relationships in face encoding. *Proceedings of the National Academy of Sciences, USA*, *106*, 5353–5358.

Insel, T. R. (2014). The NIMH research domain criteria (RDoC) project: Precision medicine for psychiatry. *American Journal of Psychiatry*, *171*, 395–397.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, *167*, 748–751.

Jack, R. E., Garrod, O. G., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology*, *24*, 187–192.

Jia, H., & Martinez, A. M. (2009). Support vector machines in face recognition with occlusions. In *IEEE Conference on Computer Vision and Pattern Recognition*, *2009* (pp. 136–141). Piscataway, NJ: IEEE.

Kemelmacher-Shlizerman, I., Seitz, S., Miller, D., & Brossard, E. (2016). The megaface benchmark: 1 million faces for recognition at scale. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway, NJ: IEEE.

Leibo, J. Z., Liao, Q., Anselmi, F., Freiwald, W. A., & Poggio, T. (2016). View-tolerant face recognition and Hebbian learning imply mirror-symmetric neural tuning to head orientation. *Current Biology*, *27*, 62–67. doi:10.1016/j.cub.2016.10.015

Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, *442*, 572–575.

Martinez, A. M. (2002). Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*, 748–763.

Martinez, A. M. (in press). Visual perception of facial expressions of emotion. *Current Opinions in Psychology*.

Martinez, A., & Du, S. (2012). A model of the perception of facial expressions of emotion by humans: Research overview and perspectives. *Journal of Machine Learning Research*, *13*, 1589–1608.

Martinez, A. M., & Kak, A. C. (2001). PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*, 228–233.

Meyers, E. M., Borzello, M., Freiwald, W. A., & Tsao, D. (2015). Intelligent information loss: The coding of facial identity, head pose, and non-face information in the macaque face patch system. *The Journal of Neuroscience*, *35*, 7069–7081.

Neth, D., & Martinez, A. M. (2009). Emotion perception in emotionless face images suggests a norm-based representation. *Journal of Vision*, *9*(1), Article 5. doi:10.1167/9.1.5

Neth, D., & Martinez, A. M. (2010). A computational shape-based model of anger and sadness justifies a configural representation of faces. *Vision Research*, *50*, 1693–1711.

Ohayon, S., Freiwald, W. A., & Tsao, D. Y. (2012). What makes a cell face selective? The importance of contrast. *Neuron*, *74*, 567–581.

Richoz, A. R., Jack, R. E., Garrod, O. G., Schyns, P. G., & Caldara, R. (2015). Reconstructing dynamic mental models of facial expressions in prosopagnosia reveals distinct representations for identity and expression. *Cortex*, *65*, 50–64.

Skerry, A. E., & Saxe, R. (2015). Neural representations of emotion are organized around abstract event features. *Current Biology*, *25*, 1945–1954.

Sormaz, M., Young, A. W., & Andrews, T. J. (2016). Contributions of feature shapes and surface cues to the recognition of facial expressions. *Vision Research*, *127*, 1–10.

Srinivasan, R., Golomb, J. D., & Martinez, A. M. (2016). A neural basis of facial action recognition in humans. *The Journal of Neuroscience*, *36*, 4434–4442.

Turk, M. (2013). Over twenty years of eigenfaces. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *9*(1s), Article 45. doi:10.1145/2490824

Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B., & Barrett, L. F. (2015). A Bayesian model of category-specific emotional brain responses. *PLoS Computational Biology*, *11*(4), e1004066. doi:10.1371/journal.pcbi.1004066

You, D., Benitez-Quiroz, C. F., & Martinez, A. M. (2014). Multiobjective optimization for model selection in kernel methods in regression. *IEEE Transactions on Neural Networks and Learning Systems*, *25*, 1879–1893.

Zhao, R., & Martinez, A. M. (2016). Labeled graph kernel for behavior analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*, 1640–1650.

Zhao, R., Wang, Y., Benitez-Quiroz, C. F., Liu, Y., & Martinez, A. M. (2016). Fast and precise face alignment and 3D shape reconstruction from a single 2D image. In *European Conference on Computer Vision* (pp. 590–603). New York, NY: Springer International Publishing.